AWS S3 - CSV

Use the S3 public link to read the CSV file directly into a pandas DataFrame

```
s3_link='https://s3.amazonaws.com/projex.dezyre.com/recommender-system-for-teleco
m-products/materials/Telecom_data.csv'
df = pd.read_csv(s3_link)
```

Database

Most organizations store their data in databases such as MS SQL. Microsoft SQL Server (MS SQL) is a relational database management system developed by Microsoft. A BAK file in Microsoft SQL Server is a backup file that contains a copy of a SQL Server database at a specific point in time. It is essentially a binary representation of the database and includes all its data, tables, schema, indexes, stored procedures, and other objects.

Installing MS SQL Management Studio

To install Microsoft SQL Server Management Studio, you can follow these steps:

1. Go to the Microsoft SQL Server Downloads page (https://www.microsoft.com/en-us/sql-server/sql-server-downloads) and click on the "Download now" button for the version of SQL Server Management Studio that you want to install.
2. Follow the instructions on the screen to download the installation file to your computer.
3. Once the download is complete, locate the installation file and double-click on it to start the installation process.

Restore a BAK file in MS SQL

1. Open SQL Server Management Studio and connect to the SQL Server instance to which you want to upload the BAK file.
2. Right-click on the Databases folder in the Object Explorer pane and select "Restore Database..." from the context menu.
3. In the "Restore Database" dialog box, select the "Device" option under the "Source" section.
4. Click on the "..." button to open the "Select backup devices" dialog box.
5. In the "Select backup devices" dialog box, click on the "Add" button to add the BAK file that you want to upload.
6. In the "Locate Backup File" dialog box, browse to the location where the BAK file is stored in the project directory under the 'data' folder, select the file, and click on the "OK" button.

7. Back in the "Select backup devices" dialog box, the BAK file you added should now be listed under "Backup media:".
8. Click on the "OK" button to close the "Select backup devices" dialog box.
9. In the "Restore Database" dialog box, you should see the BAK file listed in the "Backup sets to restore" section.
10. By default, the original database name and file locations from the BAK file will be used. If you want to restore the database with a different name or to a different location, you can modify the "To database" and "Restore as" options under the "General" section.
11. Click the "Options" tab for additional restore options.
12. If you want to overwrite an existing database with the same name, you can select the "Overwrite the existing database (WITH REPLACE)" option under the "Restore options" section.
13. Click on the "OK" button to start the restore process.
14. Once the restore process is complete, you should see a message indicating that the restore was successful.

Read Data from DB to Python

The data can be accessed by secret credentials, which will be in the following format.

```python
import pyodbc
import pandas as pd
connection = pyodbc.connect('DRIVER={ODBC Driver 17 for SQL Server};\
                    SERVER=server_name;\
                    PORT=1433;\
                    DATABASE=database_name;\
                    UID=admin;\
                    PWD=password')
```

Steps to install ODBC driver

1. Go to the Microsoft Download Center page for the ODBC Driver 17 for SQL Server: https://www.microsoft.com/en-us/download/details.aspx?id=56567
2. Select the download button that corresponds to the operating system you are using.
3. Select the language you want to use for the installer, then click the download button.
4. Once the download is complete, run the installer.
5. Accept the license terms, then select the features you want to install.
6. Choose a location to install the driver, or use the default location.
7. Complete the installation process by following the instructions provided by the installer.
8. Once the installation is complete, you can use the ODBC Driver 17 for SQL Server to connect to SQL Server databases from applications that support ODBC connectivity.

Query to read the data into Pandas

```python
query = ''''''select * from Processed_month_1_data
         UNION ALL
          select * from Processed_month_2_data
          UNION ALL
          select * from Processed_month_3_data
          UNION ALL
          select * from Processed_month_4_data
          UNION ALL
          select * from Processed_month_5_data
          UNION ALL
          select * from Processed_month_6_data
          UNION ALL
          select * from Processed_month_7_data
          UNION ALL
          select * from Processed_month_8_data
          UNION ALL
          select * from Processed_month_9_data
          UNION ALL
          select * from Processed_month_10_data
          UNION ALL
          select * from Processed_month_11_data
          UNION ALL
          select * from Processed_month_12_data
          UNION ALL
          select * from Processed_month_13_data
          UNION ALL
          select * from Processed_month_14_data'''

processed_data = pd.read_sql(query,connection)
processed_data.head()
```